SUBJECT: Comments on Bengali Yaphala
BY: Paul Nelson, Microsoft Corporation

I have circulated a proposal to come to a consensus on a standardized shaping behavior for the Bengali Yaphala. The proposal was posted on the Indic Unicode list, and shown to various software vendors and government representatives of Bangladesh and India.

There was an agreement that the method that I was proposing for the encoding of the Yaphala was acceptable and was the most logical manner in which to resolve the ambiguity of shaping RA + VIRAMA + YA. Thus, the following should be considered for the editorial committee to add to the Unicode documents on the behavior of the Bengali script.

Reph and Yaphala

Reph: The formation of the Reph form is defined in the Unicode Book, Section 9.1, Rules for Rendering, R2. Basically, the Reph is formed when a Ra which has the inherent vowel killed by the virama/halant begins a syllable. This is shown in the following example.

র + ◌্ + ম ⟶ র্ম as in কর্ম (karma)

Yaphala: The Yaphala is a post-base form of Ya and I formed when the Ya is the final consonant of a syllable cluster. In this case, the previous consonant retains is base shape and the virama/halant is combined with the following Ya. This is shown in the following example.

ক + ◌্ + য ⟶ ক্য as in বাক্য (bakyô)

Issue: An ambiguous situation is encountered when the combination of Ra + virama/halant + Ya is encountered.

র + ◌্ + য ⟶ র্য or র‍্য

Normative behavior: To resolve the ambiguity with this combination and to have consistent behavior, we need to look at the processing order of the Bengali script. When parsing the text, the ability to form the Reph is identified first and therefore the Reph form should have priority in processing. Thus, it is necessary to insert a ZWNJ character into the stream between the Ra and virama/halant to allow the virama/halant and Ya to be grouped together during processing.

র + ◌্ + য ⟶ র্য
(U+09B0 + U+09CD + U+09AF)

র + ZWNJ + ◌্ + য ⟶ র‍্য
(U+09B0 + U+200C + U+09CD + U+09AF)

In the example above, the ZWNJ is used because we are saying that we want two characters that would join by default to remain as separate entities. In cases other than where the RA is the first character in the cluster the ZWNJ is not required for the formation of the Yaphala. However, for ease of placing the Yaphala input as a single key input, it should be permissible for the Yaphala to be consistently formed by "ZWNJ + VIRAMA + YA" (U+200C + U+09CD + U+09AF).